DOCUMENT RESUME

ED 464 952 TM 033 883

AUTHOR Adams, Stephen T.

TITLE Use of a Computer Environment To Analyze the Coherence of Argumentation about Policies Proposed To Ameliorate Global Warming.

SPONS AGENCY National Science Foundation, Arlington, VA.
PUB DATE 2002-04-00
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 1-5, 2002).
CONTRACT DGE-9554564
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computer Uses in Education; Evaluation Methods; Global Warming; *High School Students; High Schools; Interviews; *Persuasive Discourse; Secondary School Teachers; Surveys
IDENTIFIERS Case Arguments; *Connectionism

ABSTRACT

In designing computer environments to support the creation of arguments, a central issue concerns evaluating the quality of these arguments. This study uses one such computer environment, Convince ME, that uniquely uses a connectionist model, ECHO, to generate Model Fit values as a kind of measure of an argument's coherence. This study sought to triangulate these Model Fit values by comparing them to other measures, including a measure of the stability of views (from public opinion researcher Daniel Yankelovich) and the number of statements in participant's arguments. Two 17-year-old students and 1 scientist participated in pre- and post-surveys and interviews and used Convince Me to create arguments about policies that would ameliorate global warming. The pattern of positive correlations among these measures, coupled with results from a debriefing interview, illuminate cognitive and educational issues involved in using Convince Me (or programs like it) to support reasoning about public policy issues. (Contains 25 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 464 952

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

S. Adams

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Use of a Computer Environment to Analyze the Coherence of Argumentation About Policies Proposed to Ameliorate Global Warming

Paper Presented at the Annual Meeting of the
American Educational Research Association
New Orleans

April 1-5, 2002

Session 50.34 Thinking, Problem Solving, and Argumentation
in Technology Environments

Stephen T. Adams

College of Education
California State University, Long Beach
1250 Bellflower Blvd.
Long Beach, CA 90840
sadams2@csulb.edu

TM033883

BEST COPY AVAILABLE

Abstract

In designing computer environments to support the creation of arguments, a central issue concerns evaluating the quality of those arguments. This study uses one such computer environment, Convince Me, which uniquely employs a connectionist model, ECHO, to generate "Model Fit" values as a kind of measure of an argument's coherence. This study seeks to "triangulate" these Model Fit values by comparing them to other measures including a measure of the stability of views (from public opinion researcher Daniel Yankelovich) and the number of statements in participant's arguments. Ten 17-year-old students and one scientist participated in pre-and post-surveys and interviews and used Convince Me to create arguments about policies that would ameliorate global warming. The pattern of positive correlations among the aforementioned measures, coupled with results from a debriefing interview, illuminate cognitive and educational issues involved in using Convince Me (or programs like it) to support reasoning about public policy issues.

Use of a Computer Environment to Analyze the Coherence of Argumentation about Policies Proposed to Ameliorate Global Warming

An emerging area of educational technology research and development involves the design of computer environments that enable students to create and evaluate arguments on specified topics. Such environments include Belvedere (Suthers, Connelly, Lesgold, Paolucci, Toth, Toth, & Weiner, 2001; Suthers & Weiner, 1995; Suthers, Weiner, Connelly, and Paolucci, 1995), Sensemaker (Bell, 1997, 1998) and Convince Me (Diehl, 2001; Diehl, Ranney, & Shank, 2001; Schank, Ranney, Hoadley, Diehl, & Neff, 1994). Of these, the design of Convince Me is unique in that it incorporates a connectionist computer model, ECHO (Thagard, 1989). Convince Me uses the ECHO model to compute a Model Fit rating which can be taken as a rough measure of the overall coherence of a given argument. In using such computer environments, questions arise regarding the cognitive meaning of the computer-based representations of their arguments. In the case of Convince Me in particular, questions arise regarding the meaning of the Model Fit ratings, which are a central construct to the environment and are taken as a kind of desideratum of good reasoning.

This study is oriented toward understanding the use of Convince Me as a tool for creating arguments about policy issues. In particular, the study seeks to "triangulate" the environment's Model Fit measure by comparing it to other measures, including measures of the stability of subjects' views. For further information, participants were asked in a debriefing interview to discuss how well they think their Convince Me arguments reflect their thinking. The study uses policies proposed to ameliorate global warming as a context, since global warming is such an important scientific and public policy problem.

Background

The perspective that good reasoning, by definition, is coherent and stable is seen in literature in political science, public opinion research, and cognitive science. In an influential article, political scientist Philip Converse (1964), posited the notion that holding

a coherent political ideology represents greater sophistication. In Converse's view, an ideology is coherent if a person's position on one issue predicts his or her position on another issue. Similarly, public opinion researcher Daniel Yankelovich uses coherence as an indicator of better opinion. He proposes two formal criteria for the quality of opinion about policy issues: consistency—whether the opinion is consistent with one's other beliefs, and volatility—whether the opinion is firmly held or changes (Yankelovich, 1991, p. 24).

Yankelovich, Skelly, & White (1981) developed a measure designed to gauge relatively quickly the latter of these criteria, volatility or stability. This measure, the Mushiness Index, consists of a battery of four questions. Three questions relate to hypothesized sources of stability: whether a person has a greater personal stake in an issue, has more information about an issue, or has had discussions with others about the issue. A fourth question simply asks respondents to estimate the likelihood that they will change their minds. Yankelovich et al (1981) found that the index was relatively good at predicting whether persons would change their positions about a policy issue.

Techniques from an area of research in cognitive science provide a way to measure the coherence of argumentation. Thagard (1989) developed a Theory of Explanatory Coherence, embodied in a computational model of reasoning, ECHO. ECHO has been used to model scientific controversies in several areas including the history of science (Thagard, 1989) and physics reasoning (e.g., Ranney & Thagard, 1988). In these studies, coherent argumentation according to ECHO was employed to evaluate subjects' reasoning.

Convince Me is a computer program that allows users to create representations of their own thinking using ECHO. The program also provides a way for users to reflect on their thinking using feedback from the program. A series of publications discuss studies of Convince Me (Diehl, 2001; Diehl, Ranney, & Schank, 2001; Ranney & Schank, 1995; Ranney, Schank, & Diehl, 1995; Ranney, Schank, Hoadley, & Neff, 1996; Schank, 1995; Schank, Hoadley, Dougery, Neff, & Ranney, 1993; Schank & Ranney, 1992; Schank & Ranney, 1993; Schank, Ranney, Hoadley, Diehl, & Neff, 1994).

Convince Me provides a way for a user to enter a set of statements, categorized into hypotheses and evidence. The user then creates a set of links between the statements, indicating which statements explain or contradict which others. In addition, the user enters a set of Believability ratings indicating how much he or she believes each individual statement. The program does not understand the meaning of the statements, but when the ECHO model is run, it generates values that may be thought of as its own ratings of the "believability" of the statements based on the structure of the argument and ECHO's rules. Finally, the program can report "Model Fit" values, which represent the correlation between the user's Believability ratings and ECHO's model-generated values. These Model Fit values are taken as a rough measure of coherent argumentation according to ECHO, if one assumes that the argument structure reflects what the subject really knows.

This study investigates whether coherence, as measured by ECHO, relates to stability, as measured by the Mushiness Index. In addition, the study checks whether the Mushiness Index ratings predict changes in positions about policies as Yankelovich found that they did. Finally, by asking participants to discuss their use of Convince Me after they have created arguments with it, the study probes for ways in which participants' knowledge and views of the program may have influenced their use of it.

Methods

Participants

A total of 11 participants (10 high school students and one scientist) partook in two experimental sessions totaling about 3.75 hours. The students were all seventeen years old and seniors from a single San Francisco Bay Area high school. They were drawn from science classes having students of mixed ability levels. Of the ten students, four were male and six were female. They were paid $5.75 per hour. The scientist, a 34-year-old male postdoctoral researcher at a major research university, was paid $12.50 per hour. He was included as a participant with much greater expertise about issue of global warming. (Note that in the Results section that follows, the names are fictitious but preserve gender.)

Materials

The materials included a Policy Questionnaire, a Mushiness Index questionnaire based on five of these policies, and the Convince Me program. The Policy Questionnaire, which was adapted from one developed by Public Agenda (Doble & Johnson, 1990; Doble, Richardson, & Danks, 1990), asked participants to rate their support for 15 policies that have been proposed to ameliorate global warming. The format of the questionnaire explicitly included a trade-off for each proposed policy, as illustrated in the following two policies:

- Raise the gasoline tax by $1.00 a gallon even if that would burden truckers and others who need their cars for work.
- "Fee-rebate" system. Charge people who buy cars with poor gas mileage an additional fee and use the money to give rebates to people who buy cars with good gas mileage. People who buy cars with good gas mileage would get rebates of up to $1,000 even if people who buy cars with poor gas mileage would be charged fees of up to $1,000.

As shall be discussed in further detail, subjects were both interviewed about these two policies and asked to create diagrams using Convince Me describing their views about them. In addition, subjects were interviewed about three other policies from the Policy Questionnaire: increasing the use of nuclear power, fertilizing the oceans with iron to stimulate phytoplankton and thereby sequester carbon dioxide, and another automobile-related policy which would require raising the fuel efficiency (mpg) requirements for automobiles. In this way, the study incorporated a variety of different types of policies, involving regulations, incentives, and technological and geoenginnering approaches. Policies affecting automobiles were emphasized both because of the expectation that they would be the most tangible to participants and because of the substantial contribution to global greenhouse gas output made by the use of automobiles in the United States.

The Mushiness Index questionnaire, designed to probe the stability of participants' views, consists of a battery of four questions as shown in Table 1.

Table 1. The Questions of the Mushiness Index

A. On a scale of 1 to 6, where 1 means that the issue affects you personally very little and 6 means that you really feel deeply involved in this issue, where would you place yourself?

B. On some issues people feel that they really have all the information that they need in order to form a strong opinion on that issue, while on other issues they would like to get additional information before solidifying their opinion. On a scale of 1 to 6, where 1 means that you feel you definitely need more information on the issue and 6 means that you do not feel you need to have any more information on the issue, where would you place yourself?

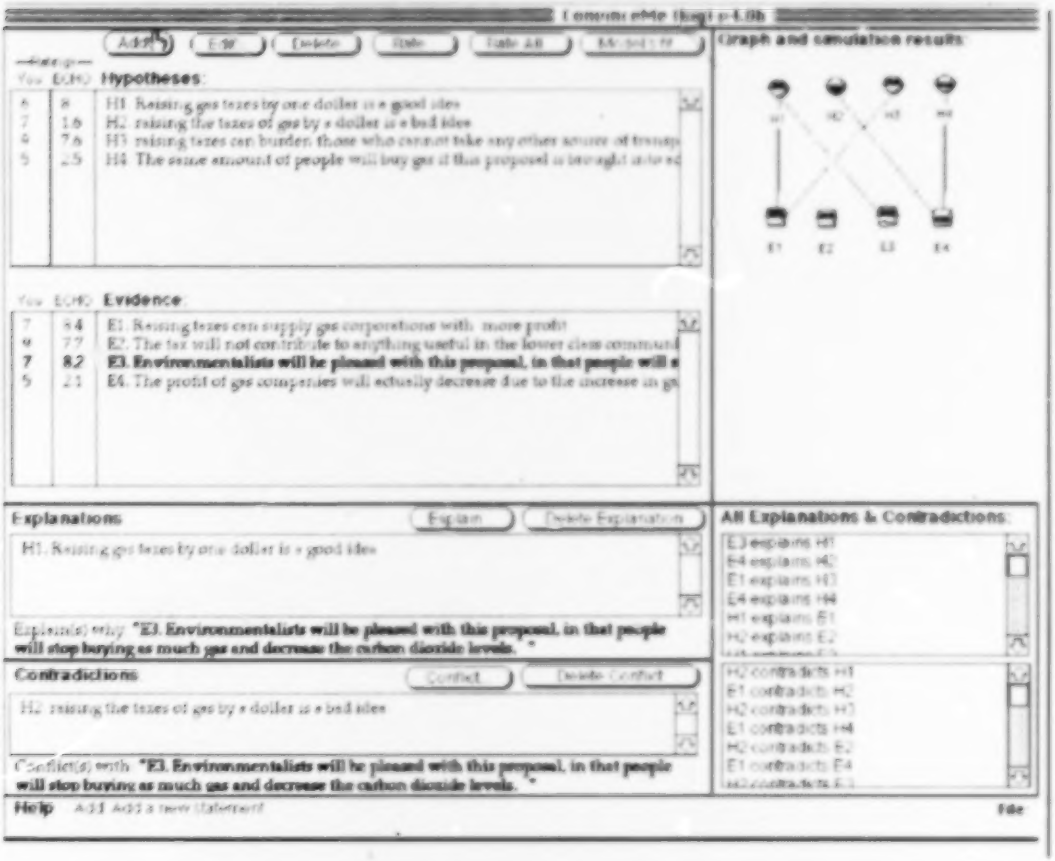
C. On the scale of 1 to 6, where 1 means that you and your friends and family rarely, if ever, discuss the issue and 6 means that you and your friends and family discuss it relatively often, where would you place yourself?

D. People have told us that on some issues they come to a conclusion and they stick with that position, no matter what. On other issues, however, they may take a position, but they know that they could change their minds pretty easily. On a scale of 1 to 6, where 1 means that you could change your mind very easily on this issue and 6 means that you are likely to stick with your position no matter what, where would you place yourself?

Note. From Yankelovich et al. (1981)

Figure 1 shows the interface of the Convince Me program.

Figure 1. The Convince Me Interface


Graph and simulation results:
Hypotheses:
H1. Raising gas taxes by one dollar is a good idea
H2. raising the taxes of gas by a dollar is a bad idea
H3. raising taxes can burden those who cannot take any other source of transp
H4. The same amount of people will buy gas if this proposed is brought into ef
Evidence:
E1. Raising taxes can supply gas corporations with more profit
E2. The tax will not contribute to anything useful in the lower class communit
E3. Environmentalists will be pleased with this proposal, in that people will s
E4. The profit of gas companies will actually decrease due to the increase in ga
Explanations
Explain
Delete Explanation
H1. Raising gas taxes by one dollar is a good idea
Explain(s) why: "E3. Environmentalists will be pleased with this proposal, in that people will stop buying as much gas and decrease the carbon dioxide levels."
All Explanations & Contradictions:
E3 explains H1
E4 explains H2
E1 explains H3
E4 explains H4
H1 explains E1
H2 explains E2
Contradictions
Conflict
Delete Conflict
H2. raising the taxes of gas by a dollar is a bad idea
Conflict(s) with: "E3. Environmentalists will be pleased with this proposal, in that people will stop buying as much gas and decrease the carbon dioxide levels."
H2 contradicts H1
E1 contradicts H2
H2 contradicts H3
E1 contradicts H4
H2 contradicts E2
E1 contradicts E4
Help
File

In the Debriefing Interview, participants were asked to respond to two questions: "What did you think about the Convince Me program?" and "How well do you think your Convince Me argument represents your knowledge and beliefs?" Participants were asked follow-up questions on a case-by-case basis.

Procedure

In Session 1, participants were introduced to the purpose of the study and given a pretest Policy Questionnaire and a pretest Mushiness Index Questionnaire. Then, in preparation for the second session, they participated in an interview and briefing concerning the science and uncertainties of global warming, some policy options proposed to ameliorate global warming, and related issues. (Content analyses were published separately (Adams, 1999, 2001.) In Session 2, participants were interviewed and created

BEST COPY AVAILABLE

Convince Me diagrams about two of the policies. For each policy, participants also took the Mushiness Index questionnaire immediately after completing their Convince Me diagram for that policy. They also took the Policy Questionnaire a second time after completing their Convince Me diagrams. Finally, participants were given a Debriefing Interview regarding their experiences and reflections about the experimental session.

Administering the questionnaire at the beginning and at the end of the experimental sessions provided a way to check for any changes in subjects' positions. For two of the policies, concerning the gasoline tax and fee-rebate system, subjects also created Convince Me diagrams on the computer. These policies were selected since they both involved the common theme of the automobile, while potentially differing in terms of the degree to which subjects were likely to have firm opinions about them. Specifically, it was expected that subjects would have firmer opinions about the gasoline tax proposal than the feebate proposal. The order of these Convince Me exercises was counterbalanced across the subjects.

As an indicator of the stability of subjects' support for the five policies, the Mushiness Index was administered to participants at the beginning of the experimental session. It was hypothesized that the experimental interventions, including the briefing and interviews about the policies, could possibly increase subjects' overall knowledge about the issues. Also, by providing a setting for participants to think through the implications of adopting or not adopting the policies, the intervention could possibly prompt participants to change their positions. Therefore, the Mushiness Index was administered a second time after the interviews and after participants had finished their Convince Me diagrams. This second Mushiness Index questionnaire was conducted immediately before the Model Fit value was taken so that the two measures would both reflect students' cognitive state following the interview and Convince Me exercise. In other words, the second Mushiness Index score provides a temporally critical comparison to the Model Fit value. On the other hand, the first Mushiness Index score, given before the interviews, was taken to determine whether the Mushiness Index predicts changes in the level of support for a policy.

Results

Subjects' support for policies increased somewhat from the pretest and posttest, and firmer ratings on the Mushiness Index corresponded to fewer position changes—results consistent with Yankelovich's finding that Mushiness Index ratings predict the stability of opinions about policy issues. Further, positive correlations were found among ECHO Model Fit values, Mushiness Index ratings, and the number of statements in a Convince Me diagram.

Overall, the average pretest Mushiness Index score for all five policies was 11.6 (S.D. = 4.6) on a scale where 4 represents the least firm beliefs and 24 represents the most firm beliefs. In that 11 subjects were interviewed and took the Mushiness Index questionnaire for five policies, there were a total of 55 opportunities for subjects to change their positions about their support for a policy. There were sixteen instances of position changes. The average pretest Mushiness Index score for those policies with a position change, 9.7, was significantly less firm than when there was no change in position, 12.4, $t(53) = 2.01$, $p < .05$). In other words, firmer ratings on the Mushiness Index corresponded to fewer position changes.

Overall, support for policies tended to increase somewhat from the pretest to the posttest. Support for policies was coded numerically, where opposing the actions of a policy corresponded to -1, supporting a policy corresponded to +1, and being not sure corresponded to 0. On the pretest, the mean overall support for the five policies was -0.23 (somewhat opposed), with a standard deviation of 0.92. By the posttest, the average support increased to +0.36 (somewhat supportive), with a standard deviation of 0.95. Thus, the mean difference increased by 0.59, which is significant, $t(21) = 2.89$, $p < .009$. This is consistent with the expectation that the experimental intervention would serve to heighten subjects' knowledge about global warming and their willingness to support policies to ameliorate it.

Support for the feebate policy increased dramatically while support for the gasoline tax policy did not change much. On a numeric scale where -1 denotes opposition and +1 denotes support for a policy, average support ratings for the feebate policy rose significantly, from -0.27 to 0.64, $t(10) = 3.19$, $p < 0.01$. On the other hand, average

support ratings for the gasoline tax policy were -0.18 on the pretest and 0.09 on the posttest, which was not a significant increase, $t(10) = 1.00$, $p < .34$.

Table 2 gives descriptive statistics of the Convince Me diagrams that subjects created. The average Model Fit value ($r = .57$) was similar to the average Model Fit value in Schank's (1995) study ($r = .53$). Model Fit values were somewhat higher for the feebate problem ($r = .62$) than the gasoline tax problem ($r = .52$). However, this difference was not significant, $t(20) = 0.67$, $p < 0.67$. The correlation between Model Fit values and Mushiness Index posttest scores was also higher in the feebate problem ($r = .63$, $p < .04$) than the gasoline tax problem ($r = .51$, $p < .11$). One Model Fit value, for a student's gasoline tax problem, was anomalously low (-0.32) and will be discussed further in the "Debriefing Interview" section.

Table 2. Descriptive Statistics of Convince Me Diagrams

	mean			s.d			minimum			maximum		
	Fee	tax	both	fee	tax	both	fee	tax	both	fee	Tax	both
MF[a]	.62	.52	.57	.36	.37	.36	-.11	-.32	-.32	.98	.88	.98
H[b]	4.3	5.0	4.6	2.6	3.4	3	2	2	2	10	12	12
E[c]	5.0	5.2	5.1	2.1	2.2	2.1	1	1	1	8	9	9
H+E	9.3	10.3	9.8	2.6	2.2	2.4	4	7	4	14	13	14
Ex[d]	6.5	7.3	6.9	3.1	3.1	3.1	2	3	2	13	15	15
C[e]	6.3	7.3	6.8	3.0	2.7	2.8	1	1	1	10	11	11
Ex + C	12.8	14.5	13.7	4.5	4.7	4.6	5	8	5	17	25	25

[a]Model Fit values. [b]Hypothesis statements. [c]Evidence statements. [d]Explanation links. [e]Contradiction links.

Model Fit Values, Mushiness Index Scores, and Number of Statements

Positive correlations were found between Model Fit values, Mushiness Index scores, and the number of statements in a Convince Me diagram. The correlation between Model Fit values and Mushiness Index posttest scores was 0.48 ($p < 0.02$), without the outlier, this increased to .55 ($p < .008$). The correlation between Model Fit values and the number of statements in a Convince Me diagram was 0.34 ($p < 0.12$), without the outlier, this rose to 0.50 ($p < .03$). The correlation between the number of statements in a diagram and Mushiness Index posttest scores was ($r = 0.45$, $p < 0.04$), without the outlier, the

statistic was the same. Furthermore, pretest Mushiness Index scores were also positively and significantly correlated with the number of statements in a Convince Me diagram (r = 0.56, p < 0.007), and the correlation was similar without the outlier (r = 0.56, p < 0.008).

There was a positive correlation between posttest Mushiness Index scores and Model Fit values (r = .483, p < .02). However, pretest Mushiness Index scores were not significantly correlated with Model Fit values (r = .207, p < .36). A z test of whether these correlations differed significantly (i.e., .207 vs. .483) was suggestive but not conclusive (z = 1.38, p < .17).

The positive correlation between Model Fit values and Mushiness Index ratings supports the theoretical expectation that stability and coherence are related. This raises an additional question of whether Model Fit ratings correlate more closely to some questions of the Mushiness Index than to others. Table 3 shows a correlation matrix of the individual Mushiness Index items. See Table 1 for the full text of the questions, which correspond to (a) personal involvement, (b) amount of information, (c) discussing an issue with friends and family, and (d) self-assessment of the firmness of one's opinon. The sum of these four items is the overall Mushiness Index score. The item most highly correlated with Model Fit values was Question D (r = .57), concerning the likelihood of changing one's mind. The item least correlated with Model Fit values was Question B (r = .18), which concerned having sufficient information about the issue.

Table 3: Correlation Matrix of Individual Mushiness Index Items

	A	B	C	D	a+b+c+d	MF[a]
A	-	.54	.68	.57	.92	.30
B		-	.24	.37	.64	.18
C			-	.27	.80	.43
D				-	.66	.57
a+b+c+d					-	.48
MF[a]						-

[a]Model Fit values

Debriefing Interview

Data from the debriefing interviews suggest further influences on participants' construction of arguments.

Students may view some statements in a Convince Me diagram as much more important than others. As noted above, the model fit ratings for one participant's gasoline tax diagram was anomalously low: -0.32. The student's ECHO representation had six statements linked to the hypothesis he believed more strongly ("raising the gasoline tax by $1.00 is not a good idea") but only three statements linked to the opposing hypothesis that he believed less strongly. It turned out the student felt that one of the three statements ("The tax would change the whole fabric of society...") was more important than all the others. The student explained:

> *The real reason why I lean on the side of raising the gasoline tax by one dollar is not a good idea, is because it's gonna affect the way our society is made up... and that's the biggest factor. But I still agree with arguments such as "people aren't gonna drive as much cars,"... they're both true.*

This outlier shows an example of a situation that can create a problem for ECHO. ECHO does give more weight to evidential than to hypothetical statements but does not have a way to represent or weight the "importance" of a hypothesis. If a user of Convince Me does not link multiple statements to an "important" hypothesis, ECHO might tend to rate that hypothesis lower than the user did.

Students with more interest in an issue may generate more statements. As indicated earlier, the data indicate that both pretest and posttest Mushiness Index scores were correlated with the number of statements subjects wrote in a Convince Me diagram. One of the clearest examples of this pattern came from differences in how one subject responded to the gasoline tax and feebate policies. Her pretest Mushiness Index score for the gasoline tax problem (15) was nearly twice as firm as her score on that test for the

feebate problem (8). Similarly, her Convince Me diagram for the gasoline tax problem contained twice as many statements (8) as her diagram for the feebate policy (4).

In the Convince Me debriefing interview, she said that she liked using Convince Me for the gasoline tax problem: "The one about gasoline, that was fun." She indicated that it was more fun than the feebate problem because it was more relevant to her:

> *I'm not buying a car right now...so [the feebate policy] didn't really apply to me. But gas applies to me, because I drive a car, and I drive my mom's car a lot, and my grandmother's car, so it really applies to me.*

As discussed earlier, one of the questions of the Mushiness Index specifically checks for this kind of perceived personal relevance. (See the first question of Table 1). Personal relevance may contribute to a subject's motivation and interest in creating Convince Me diagrams.

Differing stances of users toward Convince Me may influence the use of the program. Participants differed in their views of using Convince Me, as illustrated by the contrasting responses of the students Marie, who was positive about the program, and Howard, who was critical of it. Marie indicated that Convince Me helped her clarify her opinion on the feebate policy. Her statement, below, was in response to a neutral prompt, "So, what did you think of using Convince Me?"

> *That was ... a good learning tool....I sort of got more into detail and I was questioning my ideas about what kinds of businesses would need low mileage trucks and cars, and the awards for buying high mileage cars....When I did this question [on the pretest questionnaire], I was more indecisive. I didn't really have a point of view on the argument, and [Convince Me] helped me to get one, whether it was good or bad...I was just sort of, I didn't*

> *really know if that's a good or a bad idea, but now I've kind of decided it's overall a good idea.*

Consistent with her explanation that she formed a clearer opinion on the policy, her Mushiness Index score increased from 7 on the pretest to 11 on the posttest. When asked to compare the experience of being interviewed with using Convince Me, she indicated that she preferred Convince Me:

> *I think this is better, at the computer, because it, you get more time to think about things and you can make more of a strong link between different arguments, and it's a lot more concrete.*

On the other hand, Howard, while acknowledging that using the program helped him clarify his thoughts, indicated that he was critical of using it:

> *S: What was the point of it? I guess the point was sort of to clarify in my mind what I thought about these things? I think it helped about as much as writing down my own thoughts would. I don't think it was really necessary...*

He viewed himself as having an adversarial relationship with the computer:

> *S: That felt like I was, I should've been trying to please the computer, or trying to trick the computer. It sort of sets up an adversarial – well, for me it sets up an adversarial relationship between how stupid the computer is and how much smarter I am than it. I just don't want to play chess against it.*

Just before he finished it, Howard's Convince Me diagram for the feebate problem contained three statements linked with a hypothesis supporting the feebate policy but only two statements linked with the hypothesis opposing the policy. Then he added a flippant hypothesis: "Oodles of green poodles jump der strudle." He linked this statement to the

hypothesis opposing the feebate proposal—the side that had one fewer statement. When asked why he included this statement, he indicated he was "balancing" the diagram. In effect, by adding one more statement to the argument with the fewer hypotheses, he improved how Convince Me's model would evaluate the believability of the statements.

Howard was also candid about his perception that his Convince Me arguments were limited by the fact that he wasn't very knowledgeable about the issues and didn't necessarily believe what he included in his arguments. When asked how well his Convince Me diagrams reflected his thinking, Howard responded:

> *S: Not really well, but honestly I don't think about this that much, and it may reflect that. Not terribly well, but as well as could be expected I suppose. If I had known, if I had been more decisive, if I was more knowledgeable about the subjects, it might be helpful.*
>
> *E: Hm, and why would that be?*
>
> *S: Well, I was sorta making stuff up. I was sort of, okay, this is what I've heard. This isn't necessarily what I believe. I only really, I really believed what I wrote down only two or three times, and I certainly didn't believe about, what I wrote about poodles.*

Howard's remarks that he was "sorta making stuff up" reminds us that subjects may be happy to give experimenters a response, but that does not mean that their responses necessarily reflect much of an opinion (Fischhoff, Slovic, & Lichenstein, 1988).

Discussion

A central question for computer environments designed to support creating and evaluating arguments is to identify what constitutes a "better" or "worse" argument. The Model Fit from the Convince Me environment provides such a measure, which relates to the coherence of the argument structure. Consistent with the theoretical expectation that

coherence and stability would be related, this study has found positive correlations among Convince Me's Model Fit measure and Yankelovich's Mushiness Index. On the other hand, the study also identifies some of the complexities underlying the Model Fit measure. From an educational standpoint, if Model Fit values tend to increase with the firmness of one's opinion, it is worthwhile to ask whether "firmer" opinions are necessarily "better" opinions. Although if ignored, this issue could represent a problem for using Convince Me's Model Fit measure as a metric of a good argument, the dilemma can also be turned to educationally productive ends. For example, to stimulate students to reflect about characteristics of good arguments, students using Convince Me could be asked to develop criteria for determining whether and in which circumstances they view higher Model Fit values as indicative of "better" or "worse" arguments.

Through its inclusion of a debriefing interview, the study identifies further considerations when using Convince Me. The data suggest that ECHO models may be less robust in cases where the relative importance of beliefs is unevenly represented in participant's Convince Me diagrams. A way to weight the importance of statements could be incorporated into the program, or users could be encouraged to more evenly represent their beliefs. The debriefing interview also illustrates ways in which the stances that participants adopt towards using the program may influence the representations they create and the educational experiences they derive from using it. This is clearly an important part of self-directed educational activities such as this. Subsequent research could further investigate students' stances towards the program and ways of encouraging using it in a constructive manner.

Acknowledgments

This research was supported, in part, by the United States National Science Foundation (Grant No. DGE-9554564). Andrea diSessa, Michael Ranney, and Mark Christensen provided helpful suggestions for the overall research that included this study. I would also like to acknowledge members of the ECHO Educational Project of the University of California, Berkeley, including Patti Schank, Christine Diehl, Todd Shimoda, Anna Thanukos, Marcelle Siegel, Jeanne Widener, and Sarah Brem, for feedback and discussions throughout the project.

References

Adams (1999). Views of policies affecting automobiles: A comparison of high school students and specialists. *Bulletin of Science, Technology, & Society, 19*(5), 372-380.

Adams (2001). Views of the uncertainties of climate change: A comparison of high school students and specialists. *Canadian Journal of Environmental Education, 6*, 58-76.

Bell, P. (1997). *Using argument representations to make thinking visible for individuals and groups.* In R. Hall, N. Miyake, & N. Enyedy (Eds), Proceedings of CSCL '97: The Second International Confererence on Computer Support for Cooperative Learning (pp. 10-19). Toronto: University of Toronto Press.

Bell, P. (1998). The KIE software and curriculum: Relating debate activities and conceptual change through design experiments. Paper presented at the Annual Conference of the American Educational Research Association 1998, San Diego, CA.

Converse, P. E. (1964). The nature of belief systems in mass publics. In D. Apter (Ed.), *Ideology and Discontent* (pp. 206-61). New York, NY: Free Press.

Diehl, C. (2001). Computers and Students as Instructional Partners: The Role of Simulation Feedback in Collaborative Argumentation. Unpublished doctoral dissertation, University of California, Berkeley.

Diehl, C., Ranney, M., & Schank, P. (2001, March). Model-Based Feedback Supports Reflective Activity in Collaborative Argumentation. *Proceedings for the European Conference on Computer-Supported Collaborative Learning*, Maastricht, The Netherlands.

Doble, J., & Johnson, J. (1990). Science and the public: A report in three volumes: Volume I: Searching for common ground on issues related to science and technology. New York, NY: Public Agenda Foundation.

Doble, J., Richardson, A., & Danks, A. (1990). *Science and the public: A report in three volumes. Volume III: Global warming caused by the greenhouse effect*. New York: Public Agenda Foundation.

Ranney, M., & Schank, P. (1995). Protocol modeling, bifurcation/bootstrapping, and Convince Me: Computer-based methods for studying beliefs and their revision. *Behavior Research Methods, Instruments and Computers, 27*, 239-243.

Ranney, M., & Thagard, P. (1988). *Explanatory coherence and belief revision in naive physics.* Paper presented at the Tenth Annual Conference of the Cognitive Science Society.

Ranney, M., Schank, P., & Diehl, C. (1995). Competence versus performance in critical reasoning: Reducing the gap by using Convince Me. *Psychology Teaching Review, 4*(2), 153-166.

Ranney, M., Schank, P., Hoadley, C., & Neff, J. (1996). "I know one when I see one": How (much) do hypotheses differ from evidence? In R. Fidel, B. H. Kwasnik, C. Beghtol, & P. J. Smith (Eds.) Advances in classification research: Vol. 5. (ASIS Monograph Series, pp. 141-158, etc.) Medford, NJ: Learned Information.

Ranney, M., Schank, P., Hoadley, C., & Neff, J. (1996). "I know one when I see one": How (much) do hypotheses differ from evidence? In R. Fidel, B. H. Kwasnik, C. Beghtol, & P. J. Smith (Eds.) Advances in classification research: Vol. 5. (ASIS Monograph Series, pp. 141-158, etc.) Medford, NJ: Learned Information.

Schank, P. K. (1995). Computational Tools for Modeling and Aiding Reasoning: Assessing and Applying the Theory of Explanatory Coherence. Unpublished Doctoral Dissertation, University of California, Berkeley.

Schank, P., & Ranney, M. (1992). *Assessing explanatory coherence: A new method for integrating verbal data with models of on-line belief revision.* Paper presented at the Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society, Hillsdale, NJ.

Schank, P., & Ranney, M. (1993). Can reasoning be taught? *Educator, 7*(1), 16-21.

Schank, P., Hoadley, C., Dougery, K., Neff, J., & Ranney, M. (1993). The ECHO Educational Program (EEP) Coherent Reasoning Curriculum for Convince Me. Berkeley, CA: University of California, Berkeley, Graduate School of Education.

Schank, P., Ranney, M., Hoadley, C., Diehl, C., & Neff, J. (1994). *A reasoner's workbench for improving scientific thinking: Assessing Convince Me.* Paper presented at the Proceedings of the International Symposium of Mathematics/Science Education and Technology, Charlottesville, VA.

Suthers, D., & Weiner, A. (1995). *Groupware for developing critical discussion skills.* In Proceedings of CSCL '95: The International Conference for Computer Support for Cooperative Learning [On-line]. Available: http://www-cscl95.indiana.edu/cscl95/toc.html

Suthers, D., Weiner, A., Connelly, J., & Paolucci, M. (1995). *Belvedere: Engaging students in critical discussion of science and public policy issues.* Paper presented at the AI-Ed 95, The 7th World Conference on Artificial Intelligence in Education.

Suthers, D., Connelly, J., Lesgold, A., Paolucci, M., Toth, E. E., Toth, J., & Weiner, A. (2001). Representational and advisory guidance for students learning scientific inquiry. In K. D. Forbus and P. J. Feltovich (Eds.), *Smart machines in education:*

The coming revolution in educational technology. (pp. 7-35). Cambridge, MA: The MIT Press.

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences, 12*, 435-502.

Yankelovich, D. (1991). *Coming to public judgment: Making democracy work in a complex world*. Syracuse, NY: Syracuse University Press.

Yankelovich, Skelly, & White Inc. (1981). *The mushiness index: A refinement in public policy polling techniques*. Yankelovich, Skelly, and White, Incorporated.